DOCUMENT RESUME

ED 164 601 TM 008 158

AUTHOR Estes, Gary D.; Anderson, Judith I.
TITLE Observed Treatment Effects with Special Regression Evaluation Models in Groups with No Treatment.
PUB DATE Mar 78
NOTE 13p.; Paper presented at the Annual Meeting of the American Educational Research Association (62nd, Toronto, Ontario, Canada, March 27-31, 1978)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Achievement Gains; Arithmetic; Compensatory Education Programs; Control Groups; *Evaluation Methods; Experimental Groups; Grade 9; *Hypothesis Testing; *Mathematical Models; *Program Effectiveness; Program Evaluation; *Research Design; Secondary Education; *Tests of Significance
IDENTIFIERS Elementary Secondary Education Act Title I; *Regression Discontinuity Model; RMC Models

ABSTRACT

An empirical study was conducted in order to obtain treatment effect estimates with the Special Regression model for groups in which there was no treatment. General mathematics test scores were obtained from 730 ninth graders in city schools somewhat similar to Title I schools, but in which no special treatments were given. Hypothetical experimental and control groups were formed on the basis of pretest scores and the data were analyzed according to the RMC Model C (special regression model). Conclusions were that: (1) even minor ceiling or floor effects resulted in nonhomogeneous regression lines for treatment and comparison groups; (2) tests of statistical significance should preceed conclusions of educational significance at the local school or district level; and (3) guidelines to assist with interpreting evaluation results are needed. (Author/CTM)

Observed Treatment Effects with Special Regression Evaluation Models in Groups with No Treatment[1]

Gary D. Estes
Northwest Regional Educational Laboratory

Judith I. Anderson
The BDM Corporation

[1]Paper presented at the Annual Meeting of the American Educational Research Association
Toronto, March, 1978
Session 32.11

Abstract

An empirical study was conducted in order to obtain treatment effect estimates with the Special Regression model for groups in which there was no treatment. Results for the Special Regression Title-I evaluation model were calculated for approximately 730 ninth graders enrolled in general math classes in a large city school system. Conclusions were that: (1) even minor ceiling or floor effects resulted in nonhomogenous regression lines for treatment and comparison groups, (2) tests of statistical significance should preceed conclusions of educational significance at the local school or district level, and (3) guidelines to assist with interpreting evaluation results are needed.

Observed Treatment Effects with Special Regression Evaluation Models in Groups with No Treatment

Gary D. Estes
Northwest Regional Educational Laboratory

Judith I. Anderson
The BDM Corporation

The Elementary and Secondary Education Act of 1965 initiated federal aid to education on a large scale. The amount of ESEA Title I funds has steadily increased since 1965 and concurrently the number and type of compensatory programs supported with ESEA Title I funds have also increased. Evaluations of these projects have often been poorly designed and implemented; even when this has not been the case, comparison across projects has been difficult due to the wide variety of evaluation designs and testing instruments employed. In an effort to assess the impact of Title I programs nationwide in a consistent and meaningful manner, the U.S. Office of Education has provided districts with three basic evaluation designs.

The impact of a Title I project is defined by the three evaluation designs as the difference between the Title I students' observed performance and an estimate of what their performance would have been in the absence of the Title I program. The three evaluation designs provided are: Model A, the Norm-Referenced Model; Model B, the Control Group Model; and Model C, the Special Regression Model. This study will provide information on the Special Regression evaluation model. The specific objectives are (1) to estimate treatment effect when the Special Regression model is used with a group in which there was no treatment, i.e., does the Special Regression model provide an estimate of zero treatment effects when applied to actual data in which there was no treatment?, and (2) to provide recommendations which might be used in interpreting the results obtained with the Special Regression model in particular and all models in general.

The Special Regression model actually makes use of two regression models, the Regression Projection model and the Regression Discontinuity model. The designs are useful when all students below a designated cut-off score are assigned to the Title I program, i.e., the treatment group, and no student at or above the cut-off score is assigned to the Title I program, i.e., the comparison group. A strength of the Special Regression model is that it provides a sounder basis for establishing no-treatment posttest expectations than national norms (Horst, Tallmadge & Wood, 1975). A weakness of the Special Regression model is the assumption that a single regression line will fit the treatment and comparison groups under no-treatment conditions. The Regression Projection model also involves the assumption that a linear relationship exists between the pretest and posttest under no-treatment conditions. However, the problem of curvilinear relationships can be eliminated by using higher order regression equations (Horst, et al, 1975). An additional problem is that there may be no difference in regression line intercepts when lowest scoring Title I students make the largest gains--thereby masking a significant overall treatment effect. Procedures outlined by Ward and Jennings (1973, p. 128) can be used to process data in which there might be an interaction present.

Method

Data Source. Approximately 730 ninth grade students enrolled in general mathematics classes in schools somewhat similar to the District's Title I schools served as subjects in this study. Three tests of basic mathematical skills were administered to the students in October and May. The three tests are described below.

The Comprehensive Tests of Basic Skills (CTBS)Math Computation subtest which measures basic arithmetic skills, Form S, Level 4 was given. The test contains 48 multiple choice items and takes 40 minutes to administer.

The Shaw-Hiehle Individualized Computational Skills Test is a 60 item untimed test covering addition, subtraction, multiplication and division of whole numbers, fractions and decimals, as well as word problems. The test is not multiple choice, that is, students must write in the correct answers.

The Minimal Mathematics Proficiency Test (MMPT) was developed by the school district and consists of 64 multiple choice items covering basic arithmetic skills. The items are grouped into 16 areas with four items per area, and the student receives one point for each area passed. Once a student passes an area by correctly answering at least three of the four items in that area, he or she is not tested with those items again, i.e., at posttesting the student receives only items in those areas missed on the pretest.

Of the 730 students, 316 had pre-and posttest scores on the CTBS Math Computations subtest, 229 had pre-and posttest scores in the Shaw-Hiehle, and 473 had pre- and posttest scores on the MMPT. Cut-off scores which were comparable to those used to assign students to ESEA Title I math classes were identified for the three tests; these are included in Tables 1-3. Students with pretest scores at or below the cut-off scores were labeled the "treatment group" and students with pretest scores above the cut-off scores served as the "comparison group".

Analyses. Procedures outlined by Tallmadge and Wood (1976) were used to calculate program impact. Both the Regression Projection model, where treatment effect is taken to be the difference between the actual and estimated mean scores and

Table 1

Summary of Results on the CTBS Math

		Group	
		Treatment	Control
Number with Pre- and Posttest Scores		130	186
Pretest	Mean	416.0	521.6
	S. D.	34.4	39.8
Posttest	Mean	471.9	541.0
	S. D.	61.1	64.5
Pre-Post Correlation		.26	.55
Slope		.46	.89
Cut-off Score		457	
Combined Groups Pre-Post Correlation		.62	

Table 2

Summary of Results on the Shaw-Hiehle

		Group	
		Treatment	Control
Number with Pre- and Posttest Scores		111	118
Pretest	Mean	22.6	38.4
	S. D.	5.78	5.22
Posttest	Mean	29.0	39.2
	S. D.	7.49	7.98
Pre-Post Correlation		.540	.506
Slope		.700	.774
Cut-off Score		30	
Combined Groups Pre-Post Correlation		.70	

Table 3

Summary of Results on the MMPT

		Group	
		Treatment	Control
Number with Pre- and Posttest Scores		105	368
Pretest	Mean	3.44	10.96
	S. D.	1.29	3.20
Posttest	Mean	10.15	15.26
	S. D.	3.86	1.61
Pre-Post Correlation		.260	.453
Slope		.777	.228
Cut-off Score		5	
Combined Groups Pre-Post Correlation		.66	

Table 4

Summary of Results of Significance Tests

Measurement Instrument	Type of Analysis			
	Regression Projection		Regression Discontinuity	
	NCE Gain at Mean	t	NCE Gain at Cut-off	t
CTBS	4.40	3.14*	1.30	0.75
Shaw-Hiehle	2.15	1.78*	2.90	6.57*
MMPT	-6.30	5.95*	-4.42	6.22*

*Significant at the .05 level.

the Regression Discontinuity model, where the difference between the treatment group and the comparison group regression lines is tested at the point where they intersect the pretest cut-off score, were used. Results were presented in Normal Curve Equivalents. In addition, the significance tests proposed by Horst, Tallmadge & Wood (1975) were calculated, (See Table 4).

Results and Conclusions

Estimates of treatment impact at the pretest mean and cut-off score, respectively, were 4.4 and 1.3 Normal Curve Equivalents (NCE's)for the CTBS, -6.30 and -4.42 NCE's for the MMPT, and 2.15 and 2.90 NCE's for the Shaw-Hiehle. All *t* tests used to compare results for the "treatment" and "comparison" groups were significant at the .05 level except for regression discontinuity analysis on the CTBS.

A slight floor effect was observed for the "treatment group" on the CTBS pretest and a notable ceiling effect was observed for the "comparison group" on the MMPT posttest. No ceiling or floor effects were observed on the Shaw-Hiehle. Pretest-posttest correlations were moderate: CTBS, .62; Shaw-Hiehle, .70; and MMPT, .66.

On the CTBS and the MMPT, results were consistent with what would have been expected given the observed ceiling and floor effects, i.e. differences in treatment effect estimates at the pretest mean and cut-off score. However, on the Shaw-Hiehle, where no ceiling or floor effects were encountered, comparable positive gains were found using the regression projection and the regression discontinuity approaches. Tallmadge and Wood (1976, p.61) state that, "If both measures of the treatment effect show positive gains of comparable magnitudes, there can be little doubt that the treatment did, in fact, have impact."

The proposed procedures do not recommend examining results for statistical significance. Significance tests outlined by Horst et al (1975) were employed, however, and would lead one to conclude that, based on performance on the Shaw-Hiehle and the MMPT, a significant treatment effect existed, although in opposite directions, i.e., the effect was positive in the Shaw-Hiehle and negative on the MMPT. The significant negative treatment effects observed on the MMPT can easily be explained by ceiling effects and to a lesser extent the CTBS results might be a result of floor effects if one examines the slopes in Tables 1-3. Thus, these results suggest that it is possible that even minor ceiling or floor effects might have a significant effect on assessing project impact.

It is recommended that if local school districts are to meet the objective stated by Tallmadge and Wood (1976) "to provide meaningful, comparable information about Title I projects", that close attention be paid to model assumptions such as floor/ceiling effects and combined group's linearity in the absence of treatment. Further, it is recommended that additional study be made of the consequences of violating these assumptions in order to ensure that, in fact, there is a treatment in effect. In cases where ceiling and floor effects are observed, it might be advisable to consider alternatives to including all treatment or comparison students in the analyses. Specifically, if a ceiling effect is noted for the comparison group, an analysis which deleted all students scoring above a "ceiling" cut-off score might provide results which avoided ceiling effects and allowed for stronger interpretations of results. In fact, it is possible that the procedure of including all students who do not receive the Title I treatment in the comparison group might be modified to include only a sufficient number and range of students to provide an accurate estimate of the comparison group regression line. Testing Title I and all comparison students with a single test and avoiding floor effects with Title I students and ceiling effects within the comparison group might prove to be difficult. A procedure which allowed students with ceiling effects to be deleted from the evaluation analysis would probably assist to avoid the results produced on the MMPT in Table 4.

It is generally recommended that pre-posttest correlations be sufficiently high in order to compute a special regression analysis. The basis for this recommendation is to insure some degree of stability in the estimates made from the regression analyses. The restriction in range within groups which is produced when selection for group membership is based on the pretest score will automatically lower the pre-posttest correlations. Although the pre-posttest correlations for the three criteria used in this study would likely be judged adequate; it is possible that the pre-posttest correlations ranging from .26 to .55 within the treatment and control groups would have been judged inadequate. If procedures outlined by Guilford (1965, p. 341-344), were used to adjust these correlations for restriction in range, the adjusted correlations could then be examined for adequacy. Thus, it is recommended that within-group correlations in the special regression evaluation models be interpreted after restriction in range is taken into consideration.

The sample sizes in this study were reasonably large and statistically significant gains of two to three NCE's were observed on the Shaw-Hiehle for which there were no floor or ceiling effects. Some guidelines for interpreting evaluation results in addition to those already provided are needed. These guidelines might include factors such as gains observed in other projects and objectives based on estimated achievement increases as a result of Title I services, e.g., a ten percent increase in achievement. It is possible that the distributions of NCE gains will be useful in establishing "meaningful" project objectives as data are reported under the proposed Title I evaluation models. In any event, we would recommend against interpreting "any NCE gains as positive evidence of program success".

Finally, it appears that users of the Special Regression Title I evaluation model will need to be particularly sensitive to restrictions of model assumptions. It was concluded that the regression evaluation might be more applicable if procedures which deleted data causing floor or ceiling effects were adopted. Based upon the results of the Shaw-Hiehle test, it appears that guidelines in addition to tests of statistical significance and comparable impact estimates at the cut-off and pretest means will be needed in order to make meaningful interpretations of evaluation results. The comparison of gains at the cut-off point and pretest means is a simple method to assist in judging whether the treatment and comparison group regression lines are homogeneous. It might be reasonable to use actual tests of significance in assessing the homogeneity of regression lines.

In conclusion, the data in this paper provided evidence that care is needed in conducting and intrepreting evaluations using the Special Regression model proposed for Title I program evaluations. The potential usefulness of the model will be realized only if procedures and techniques are used with the model which assist to insure that the information contained in the data is teased out and not lost or misinterpreted.

References

Guilford, J. P. Fundamental statistics in psychology and education. New York: McGraw-Hill, 1965.

Horst, D. P., Tallmadge, G. K., & Wood, C. T. A practical guide for measuring project impact on student achievement. Washington, D. C.: U. S. Government Printing Office, 1975. (Stock No. 017-080-01460).

Tallmadge, G. K. & Wood, C. T. User's guide: ESEA Title I Evaluation and Reporting System. Washington, D. C.: U. S. Government Printing Office, 1976.

Ward, J. H. & Jennings, E. Introduction to linear models. Englewood Cliffs, New Jersey: Prentice-Hall, 1973.